# Causal Graphs

Statistical Modeling & Causal Inference | Oswald & Ramirez-Ruiz

# Agenda



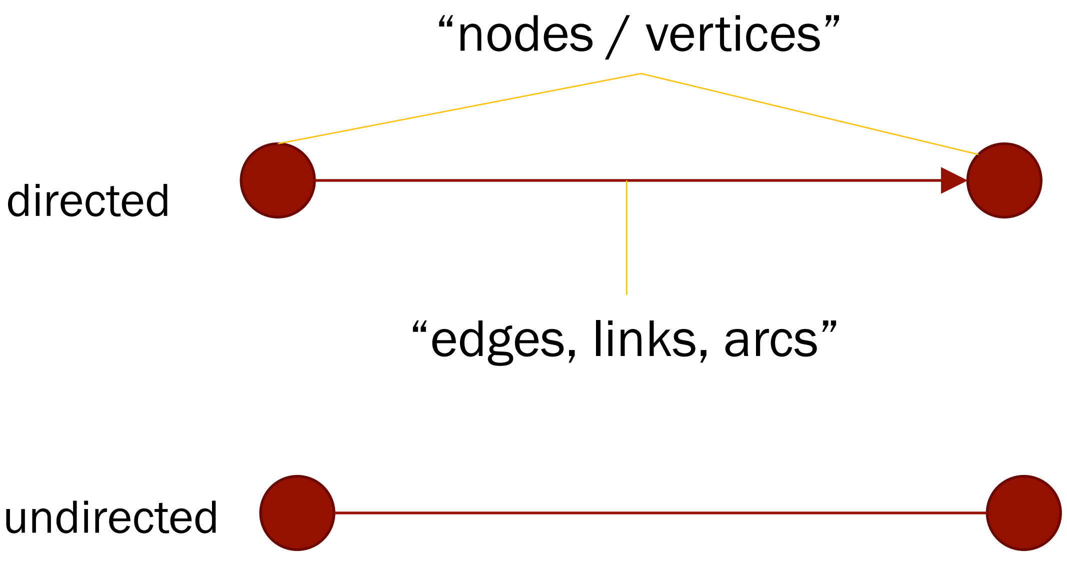
- Causal Graphs
  - DAGs
  - Translating Questions into Graphs
  - Typical Patterns

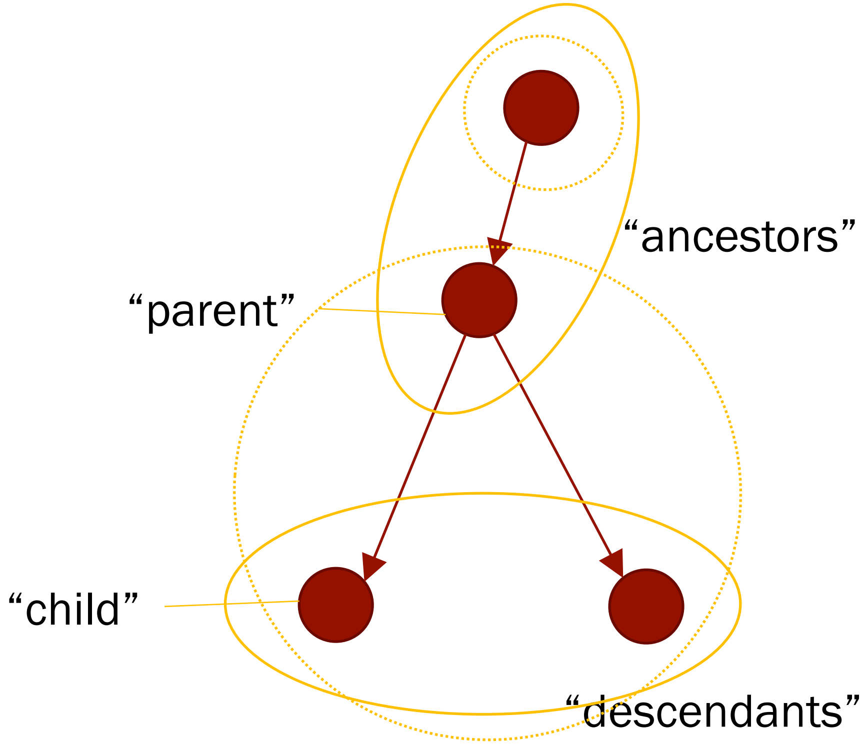
- Plotting with R
  - ggplot
  - ggdag

# Graphs



- Express beliefs about relationships among variables
- Draw conclusion about the nature of statistical associations

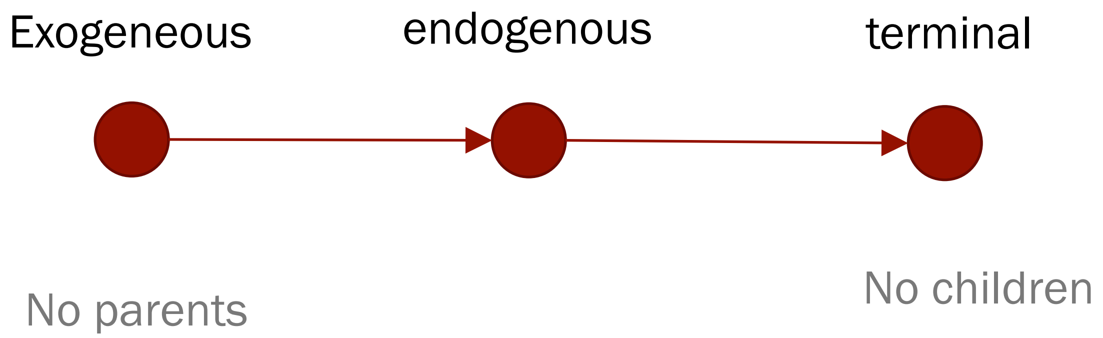


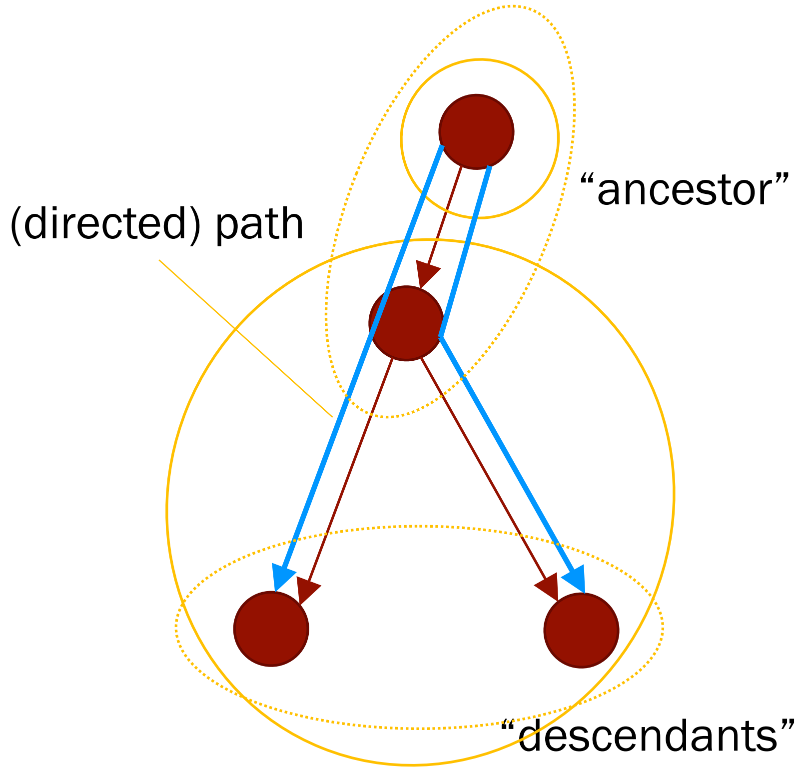



# Graphs



- Express beliefs about relationships among variables
- Draw conclusion about the nature of statistical associations

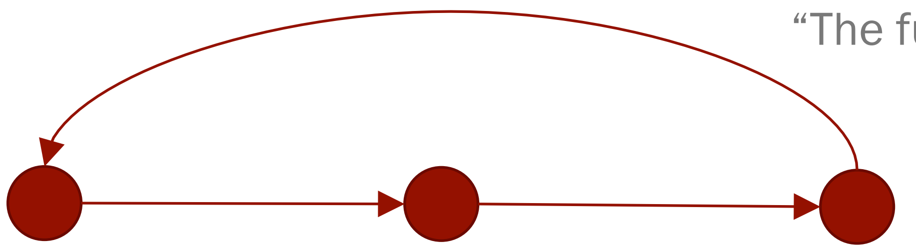


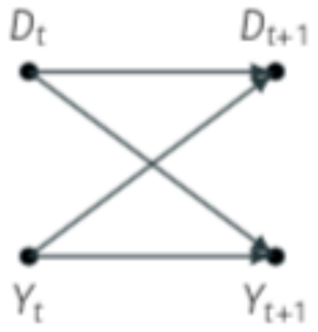
# DAGs



- "Directed acyclic graphs"
- Informal graphs expanded by adopting formal rules
- Compatible with POF but more convenient with complex causal models
- Practical for choosing "control" variables
- Encode researcher's qualitative causal assumptions
- Require theoretical and empirical knowledge



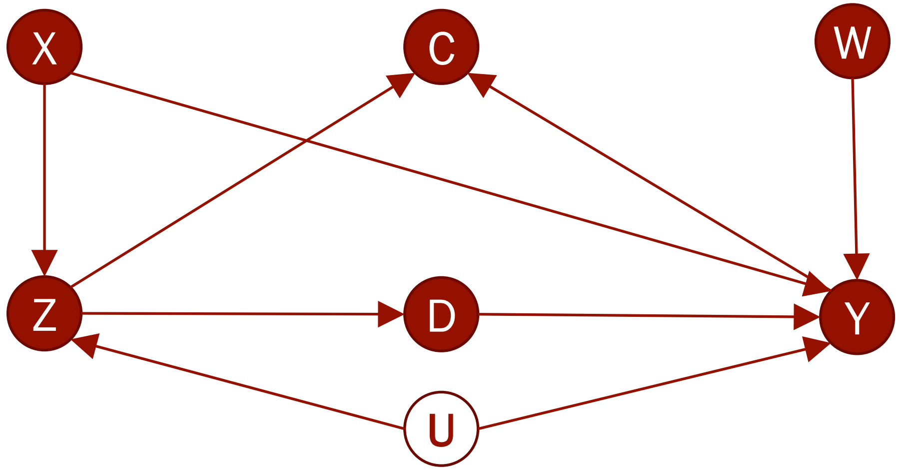
"A child cannot be a parent of their parents"
"The future cannot predict the past"



# Drawing a DAG



1. What causal relationship are you interested in? Define D & Y.
2. Collect all direct causal effects among those variables.
3. Collect all common causes of any pair of variables.
4. Also include those that you can't measure / are unobserved!
5. Cut the causes of just one variable (in case you have included them previously)

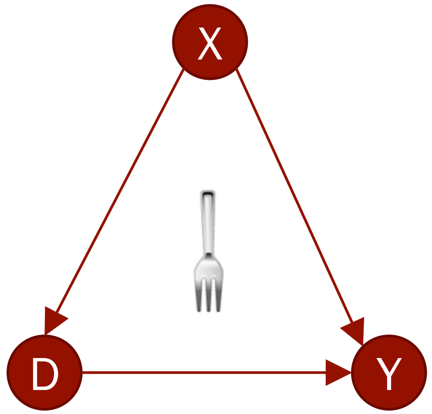



Caution: an absent causal effect is a (strong) assumption!

# DAG Patterns

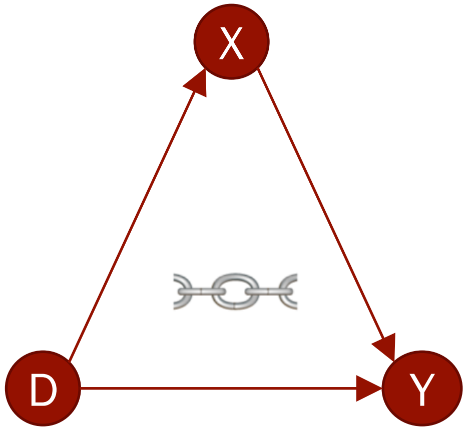


confounder



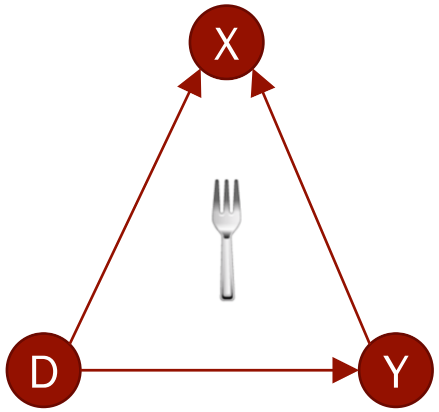

D ← X → Y
“fork”

mediator




D → X → Y
“chain”

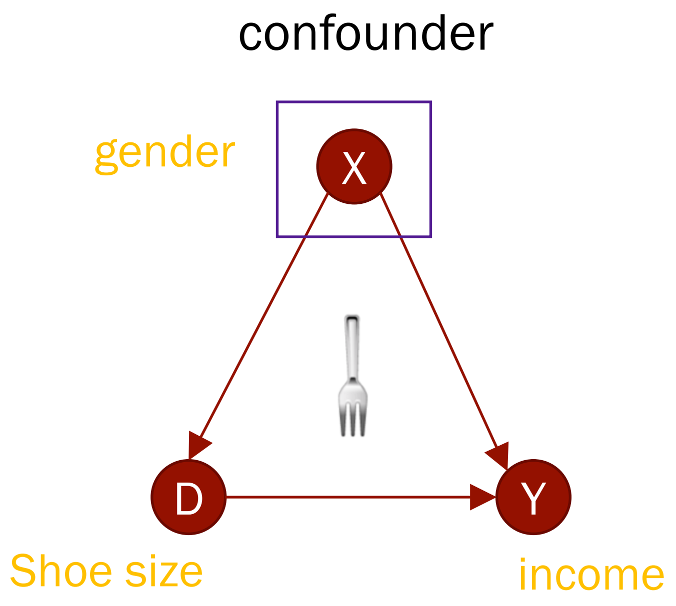
collider




D → X ← Y
“inverted fork”

# Confounders



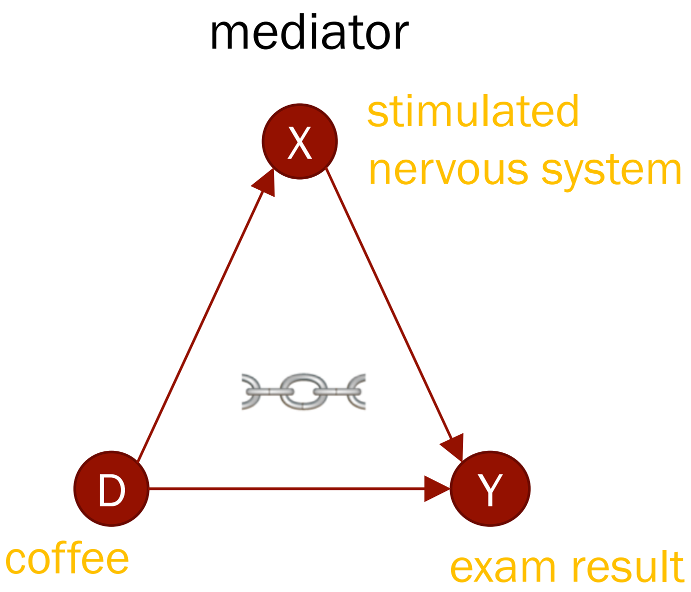



D ← X → Y
"fork"

- Induces statistical association between D and Y
- Conditioning on a confounder (or a descendant of a confounder) on the path blocks the path
- Failing to condition on confounder induces non-causal statistical association or omitted variable bias
- In most cases, you want to condition on confounders

# Mediation





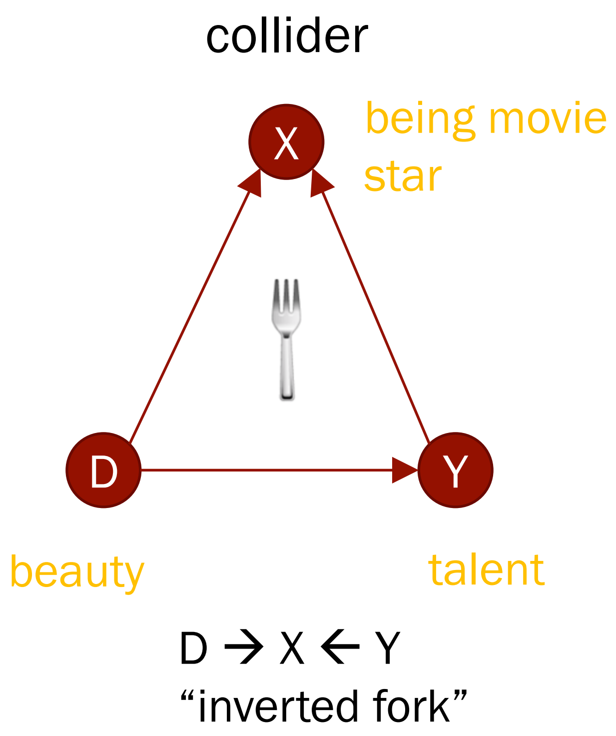

D → X → Y
"chain"

- Mediators let us express how exactly a treatment impacts the outcome → What is the mechanism?
- Cause & effect relationships can be mediated by multiple mediators
- Conditioning on them can induce bias (post-treatment bias)
- In most cases, you do <u>not</u> want to condition on mediators

# DAG Patterns



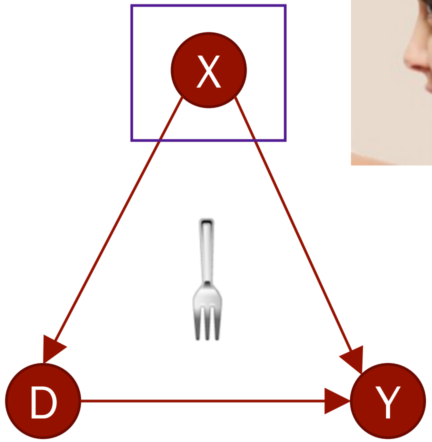
- Show how two variables jointly affect another variable
- Conditioning on collider induces statistical association between two variables (collider bias)
- In most cases, you do <u>not</u> want to condition on mediators



# Conditioning?



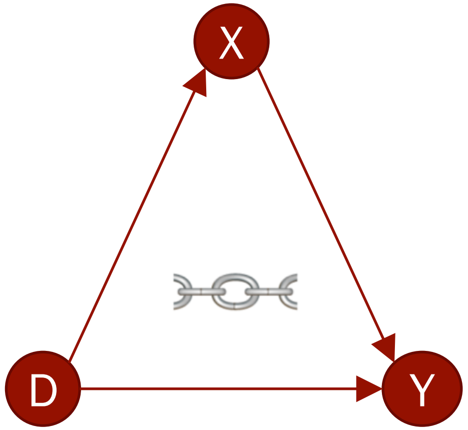
confounder





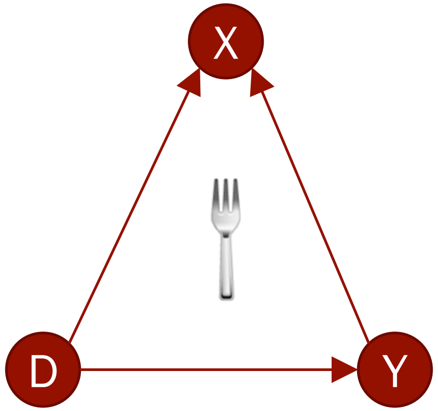
D ← X → Y
"fork"

mediator




D → X → Y
"chain"

collider




D → X ← Y
"inverted fork"

Image: https://theindianspot.com/deep-conditioning-hair-treatments/

# Plotting in R

# Further Resources



For any coding issues – [Stackoverflow]

Hertie's Data Science Lab – [Research Consulting]

Intro to ggplot – https://ggplot2-book.org/introduction.html